

# High-throughput biodiversity analysis: Rapid assessment of species richness and ecological interactions of Chrysomelidae (Coleoptera) in the tropics

Jesús Gómez-Zurita<sup>1</sup>, Anabela Cardoso<sup>1</sup>, Indiana Coronado<sup>2</sup>,  
Gissela De la Cadena<sup>1</sup>, José A. Jurado-Rivera<sup>3</sup>, Jean-Michel Maes<sup>4</sup>,  
Tinguaro Montelongo<sup>1</sup>, Dinh Thi Nguyen<sup>1,5</sup>, Anna Papadopoulou<sup>1,6</sup>

**1** *Animal Biodiversity and Evolution, Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), 08003 Barcelona, Spain* **2** *Herbario y Jardín Botánico Ambiental, Univ. Nacional Autónoma de Nicaragua, León, Nicaragua* **3** *Departament de Biologia, Universitat de les Illes Balears, 07122 Palma de Mallorca, Spain* **4** *Museo entomológico de León, León, Nicaragua* **5** *Institute of Ecology and Biological Resources, Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet, Ha Noi, Vietnam* **6** *Conservation and Evolutionary Genetics, Estación Biológica de Doñana (CSIC), 41092 Sevilla, Spain*

Corresponding author: Jesús Gómez-Zurita ([j.gomez-zurita@csic.es](mailto:j.gomez-zurita@csic.es))

---

Academic editor: J. Santiago-Blay | Received 3 November 2015 | Accepted 22 January 2016 | Published 9 June 2016

---

<http://zoobank.org/BBEFE17-74A8-48F4-8CA5-08118686EB7C>

---

**Citation:** Gómez-Zurita J, Cardoso A, Coronado I, De la Cadena G, Jurado-Rivera JA, Maes J-M, Montelongo T, Nguyen DT, Papadopoulou A (2016) High-throughput biodiversity analysis: Rapid assessment of species richness and ecological interactions of Chrysomelidae (Coleoptera) in the tropics. In: Jolivet P, Santiago-Blay J, Schmitt M (Eds) *Research on Chrysomelidae 6*. ZooKeys 597: 3–26. doi: 10.3897/zookeys.597.7065

---

## Abstract

Biodiversity assessment has been the focus of intense debate and conceptual and methodological advances in recent years. The cultural, academic and aesthetic impulses to recognise and catalogue the diversity in our surroundings, in this case of living objects, is furthermore propelled by the urgency of understanding that we may be responsible for a dramatic reduction of biodiversity, comparable in magnitude to geological mass extinctions. One of the most important advances in this attempt to characterise biodiversity has been incorporating DNA-based characters and molecular taxonomy tools to achieve faster and more efficient species delimitation and identification, even in hyperdiverse tropical biomes. In this essay we advocate for a broad understanding of Biodiversity as the inventory of species in a given environment, but also the diversity of their interactions, with both aspects being attainable using molecular markers and phylogenetic approaches. We exemplify the suitability and utility of this framework for large-scale biodiversity assessment with the results of our ongoing projects trying to characterise the communities of leaf beetles and their

host plants in several tropical setups. Moreover, we propose that approaches similar to ours, establishing the inventories of two ecologically inter-related and species-rich groups of organisms, such as insect herbivores and their angiosperm host-plants, can serve as the foundational stone to anchor a comprehensive assessment of diversity, also in tropical environments, by subsequent addition of trophic levels.

### **Keywords**

Angiosperms, Biodiversity, Chrysomelidae, insect-plant interaction, molecular ecology, molecular taxonomy, tropics

## **1. Biodiversity assessment: challenges and approaches**

### **1.1. An all-encompassing view on Biodiversity**

Few unifying concepts in Biology are so well established and ingrained in scientific and popular thinking as Biodiversity (Wilson 1988). Yet, the actual definition of Biodiversity is as encompassing and universally accepted, as it is elusive or ambiguous. Biodiversity is the diversity of Life, and by diversity of life we can understand every level of organisation, from the structural elements of genes in a particular genome, to the whole biosphere, past and present. The most intuitive idea of biodiversity has its roots in the enlightened and encyclopaedic inventorying efforts that propelled the voyages of discovery in the XVIII<sup>th</sup> Century to collect and catalogue animals, plants and minerals all over the Globe. This inventorying urge promoted in turn the creation of Museums, Zoos and Botanical Gardens in developed countries, places to keep and share with the public the records of the catalogue (institutions still reputed and alive and experimenting today a renaissance of that cataloguing spirit). Following this tradition, the word biodiversity evokes a display of life forms, or a list or catalogue of species names, ideally ranked following some system. In this context, biodiversity is tightly linked to the practice and development of Taxonomy, after all the science in charge of recognising, describing and naming organisms. Biodiversity inventories thus benefit from every conceptual and methodological advance that has contributed to the maturation of Taxonomy, from the consolidation of evolutionary thinking to the debates on species concepts, from ultrastructural analyses under the microscope to the study of gene differences among individuals or metagenomic analyses of complex environments. However, as we said, the concept of biodiversity is an all-encompassing idea that should reflect any possible way in which life is organised, including supra-organismal assemblages, such as antagonist or mutualistic associations and behaviours, food-webs, communities and biomes, their combination in ecosystems, and so on. This is essentially the diversity of ways in which life forms can interact, an aspect of biodiversity traditionally approached from Ecology, with a boost in recent times thanks to the progress made in the field of community ecology. The diversity of interactions is perhaps a less intuitive idea attached to biodiversity than the composition of a community *per se* (i.e., the idea of the inventory), but both are the complementary angles that shape the all-unifying concept of Biodiversity (Novotny and Miller 2014). Thus,

the concept of Biodiversity, certainly the one we will use throughout this essay, merges composition and functioning criteria of diversity in a given environment. These cataloguing and integrative scopes take on their highest relevance when biodiversity assessment is coupled with conservation initiatives, which ideally aim at preserving not only the nominal diversity of life forms but vitally the processes that sustain them too.

## **1.2. The challenge of Biodiversity assessment**

When the emphasis of biodiversity assessment focuses on the inventorying angle, this ‘simplified’ view on biodiversity is nonetheless generally restricted by taxonomic expertise, sampling techniques, budgetary limitations, but most of all by the sheer diversity of life forms that even the most simple biomes can harbour. A relatively homogeneous, well-delimited environment, such as a high-mountain lagoon or a monoculture crop, can be home to hundreds or thousands of different species, considering seasonality and transient and resident organisms, particularly when micro-fauna, micro-flora and, needless to say, prokaryotes are taken into account. This situation forces most biodiversity assessment plans to narrow their scope to simplified sampling strategies, e.g. canopy fogging of individual trees or deep-sea or soil probing, and typically to a specific group of organisms or habit, e.g. arthropods, insects, trees, benthic fauna, etc. Inventorying is certainly a challenge, but adding the interactions dimension to biodiversity assessment is nearly utopian. When biodiversity is described considering its functional aspects, it generally requires a much more restrictive assessment, taxonomic and for a particular interaction, e.g. pollinators of a particular plant species, community of animals exploiting a certain tree, or microorganisms with specific bioremediation potential.

These simplified approaches are defensible from an academic point of view, and they are also well adjusted to the serious underfunding for most biodiversity assessment initiatives. However, they are clearly inefficient to tackle the biological, cultural and moral problem dubbed as the Biodiversity Crisis (Western 1992; Singh 2002). Again, the challenge remains the inordinate number of species and the combinatorial of their interactions, coupled in great part with the ever-declining expertise in recognising (let alone naming) this diversity. The tip of the biodiversity crisis iceberg are groups of organisms such as angiosperms, birds, amphibians or mammals, amenable to relatively deep biodiversity assessment at least in parts of their ranges, even though the most serious concerns relate to less conspicuous but hyperdiverse groups of organisms, such as the insects (Dunn 2005). Moreover, biodiversity follows gradients, whereby the still highly unexplored tropics show the highest species counts and associations (Pianka 1966; Janzen 1973; Dowle et al. 2013), and for the known fraction of biodiversity, precisely the tropics harbour most of biological diversity perceived under significant threat, the so-called biodiversity hotspots (Myers et al. 2000). For hyperdiverse groups in hyperdiverse regions of the Planet we can generalise that our taxonomic knowledge is basic and our insight into the species ecology is merely anecdotic—although there

are, of course, important exceptions (e.g., InBIO Costa Rica; Smithsonian Institution, Barro Colorado, Panama).

The task ahead is titanic. The goal is to unravel the Earth's biodiversity as fast as possible against the ever-growing extinction rates due to habitat disappearance, fragmentation and alteration, the combined effect of climate change, overexploitation and the impact of biological invasions (Dirzo and Raven 2003; Barnosky et al. 2011). And non-trivially, the challenge is against a worrying cultural trend in this field known as the Taxonomic Impediment, the combined effect of the perception of Taxonomy as an old decaying science and the gradual disappearance of taxonomic expertise (Hebert et al. 2003; Lipscomb et al. 2003; Wheeler et al. 2004). All in all, the task is perhaps unapproachable, a mere intellectual chimera, but the scientific challenge, societal responsibility and achievable benefits are solid reasons to continue investing in biodiversity assessment, improving our assessment potential with training and technical developments.

In recent years, and as a reaction to the biodiversity crisis there has been a proliferation of initiatives aiming at large-scale biodiversity assessment. This is just to say initiatives that aim at a comprehensive (with constraints) characterisation of biodiversity, with a large regional, ecological and/or taxonomic scope. Large-scale biodiversity assessment has been a traditional practice in ecology, particularly in tropical ecology, whereby scientists sample more or less indiscriminately certain environments, providing with thousands of specimens to museums and academic laboratories around the globe. In some cases, specimens are prepared and sorted, becoming amenable for identification and cataloguing when taxonomic expertise is available. However, most typically sorting reaches a relatively high taxonomic rank, too high for meaningful community analyses, and detailed biodiversity assessment stretches indefinitely in time, depending on the interest of experts and accessibility to these collections. Today, large-scale biodiversity assessment, particularly in the context of the race against the doom to extinction of many organisms, is intimately associated to what has been referred to as rapid biodiversity assessment, in other words, quickly collecting information on the species present in a given area (Oliver and Beattie 1993; Basset et al. 1998).

### **1.3. Molecular support to biodiversity assessment**

A major boost in rapid and large-scale biodiversity assessment has been possible in the last two decades thanks to the routine implementation of molecular tools as a valuable standard to recognise diversity. The use of DNA for biodiversity assessment has provided with robust solutions for most of the challenges described above. This is a unique character system for all life forms, which is suitable for analysis with standard laboratory methods that require in turn very basic training. Thus, even modest laboratories can engage in the use of this technology for biodiversity assessment without imposing taxonomic restrictions, both in terms of scope and availability of previous knowledge, but also in terms of required taxonomic expertise (Tautz et al. 2003). Also helping



the routinely use of these approaches, the cost associated to DNA-based biodiversity assessment keeps dropping as the methods become more efficient and technology less exclusive (Yu et al. 2012). The basic laboratory steps in this procedure fundamentally require the use of DNA isolation techniques, on an individual or environmental basis (e.g., soil sample, residue of filtered sea water, ...), traditionally followed by PCR-based amplification protocols of specific genome regions, *a priori* defined standards for analysis, and finally sequencing of these markers. The process is facilitated because the latter stage can be handed to a profusion of biotechnology companies that offer sequencing services at very competitive prices. Moreover, the innovative boost of sequencing technologies of the past decade, methods collectively known as next-generation sequencing, all free from the limitations of Sanger technology, has facilitated the analysis of environmental samples and little by little displacing the need for an intermediate PCR step in some applications relevant for biodiversity assessment (Timmermans et al. 2010; Zhou et al. 2013; Andújar et al. 2015).

The use of these affordable, classical and revolutionary methodologies can potentially generate uncountable objective data for analysis, huge numbers of nucleotide characters in DNA sequences only limited by the size of the respective genomes involved, whose variability can inform of species diversity in a sample. While data can easily grow to vast amounts, these are nonetheless amenable for study even with modest computational power, given their suitability for large-scale information technology data storage and analyses. Thanks to the incorporation of molecular tools to the toolkit of taxonomists and ecologists, now the challenge and budgetary needs for biodiversity assessment are not anymore on the generation of raw data, but again on the acquisition of samples, on financing fieldwork and expeditions for biological prospection. There is still an important need for specialisation to some extent, in this case to use and develop methods to extract relevant information from collections of DNA sequences for sound biodiversity assessment. Large-scale biodiversity assessment thus rests on a new pillar as important as taxonomy and ecology: bioinformatics. The bioinformatics for biodiversity assessment has experienced an important development, receiving and exploiting the advances of more than half a century of numerical taxonomy and phylogenetics, but also the suitability of DNA sequence data for digital storage and the availability of an ever growing public database for DNA data generated worldwide.

#### **1.4. Large-scale DNA-based biodiversity assessment**

There are several ways to approach the use of DNA sequences for objective species delimitation and/or identification, but they can be divided fundamentally in two main categories. The first type of approach takes advantage of the easiness for computation of differences among DNA sequences and the assumption of a relatively uniform divergence threshold between intraspecific and interspecific DNA sequence variation. These numerical or phenetic approaches to biodiversity assessment evaluate the match of a sequence of unknown origin against comparable sequence information in a refer-

ence database (e.g., via BLAST algorithms; Altschul et al. 1997), or take advantage of more or less sophisticated clustering algorithms to facilitate taxonomic assignment. The most successful initiative following this strategy is the so-called DNA-barcoding (Hebert et al. 2003), which puts the emphasis on species identification. The second type of approaches aims at extracting evolutionary, phylogenetic information from matrices of homologous DNA-sequences to guide species inference (e.g., Wiens and Penkrot 2002). In this case, there are no implicit divergence thresholds, but there is a strong bearing on the concept of monophyly and inference of processes related to the species problem, e.g. gene flow, recombination, incomplete lineage sorting or hybridisation, among others. This field has flourished in the past few years thanks to advances in two areas of research. One is integrative taxonomy (Dayrat 2005; Schlick-Steiner 2010; Andújar et al. 2014), which tries to formalise the procedures to manage multiple sources of data, with a predominant role of molecular data, in defining (and discovering) species. The other encompasses the conceptual and methodological progress on procedures collectively known as species-trees methods, which use coalescence theory to incorporate discordance among multiple gene trees and predict species boundaries (Yang and Ranala 2010; Fujita et al. 2012). In general, phylogenetic methods have found a better use for problems related to species delimitation.

Phenetic approaches are particularly well suited for large-scale biodiversity assessment by virtue of straightforwardness and speed of analysis. However, they have some drawbacks as well. Their hypothetical optimal performance is achieved when there is a complete reference library available for comparisons (Ekrem et al. 2007), and a consistent barcoding gap or species-diagnostic behaviour of the marker of choice (Meyer and Paulay 2005). These criteria may be met for specific groups, but they are not universal. The quality and coverage of reference libraries can improve over time as new data enter the system, but there is a limitation imposed by the Taxonomic Impediment itself in providing solid taxonomies attached to the reference sequences, not to mention the fundamental problem of incompleteness of the inventory of Life. In any case, reliance on a static barcoding gap will always represent a problem, since this is not a universal, intrinsic property of species and DNA data (Meyer and Paulay 2005; Meier et al. 2006). Indeed, some alternatives exist to customise the concept of species thresholds, such as the ABGD method (Puillandre et al. 2012), but there will be always problematic groups for this criterion, e.g. species that hybridise, recent speciation events, convergence and evolutionary stasis or lineage-specific differences in evolutionary rates for the marker of choice. Moreover, taxonomic gaps in the reference library and exceptions to the barcoding gap do not prevent these approaches from producing species inferences even in the absence of true conspecifics in the reference database; these are known as false positives, and constitute one of their most serious limitations (Ross et al. 2008).

In turn, phylogenetic approaches are powerful and can assist both species delimitation and identification when used with a reference. In this case, even if the reference library does not include conspecific data, phylogenetic inference protects against false positives at the expense of taxonomic resolution (Ross et al. 2008; Berger et al. 2011).

Phylogenetic theory and practice have pushed dramatic advances in speed of analysis, both with more efficient and faster algorithms and a better use of computing capacities with parallelisation of complex calculations. However, these methodologies tend to be complex analytically, intense computationally and generally benefit from studying multiple markers, therefore are slower, less intuitive and need more training than their phenetic counterpart. Moreover, the performance of phylogenetic inference varies depending on the markers and underlying assumptions, which advises against blind attempts to conduct biodiversity assessment, without a way to evaluate systematically the robustness of the phylogenetic trees.

Clearly, DNA-based biodiversity assessment in the context of large-scale studies, can benefit of tree-based approaches taken from the field of molecular systematics, but it also requires speed of analysis. Specifically related to the problem of species identification, bacterial molecular taxonomy and current efforts to characterise microbiotas in multiple environments (e.g., Human Microbiome Project or TerraGenome) have built upon this tree-based concept for many years now. Thus, in this field, researchers exploit fast maximum likelihood phylogenetic analyses of query prokaryote 16S sequences against curated taxonomic references for this marker, e.g. workbench of Greengenes, SILVA and others (McDonald et al. 2012). Inspired by the philosophy of bacterial taxonomy, we have recently developed an analogous strategy for any kind of organism adding flexibility for the marker of choice by exploiting real-time taxonomically-tagged sequence availability in public nucleotide sequence databases [see section 2.3], the so-called BAGpipe protocol originally applied to angiosperm identification based on *psbA-trnH* data (Papadopoulou et al. 2015).

## **2. The contribution of Chrysomelidae to a diverse world**

### **2.1. Leaf beetle communities matter to large-scale biodiversity assessment**

The field of conservation biology has relied on bioindicators to monitor the quality of the environment (Noss 1990; Caro and O'Doherty 1999). Rather than attempting massive biodiversity studies on particular environments, perhaps a sound way to enhance biodiversity assessment could find inspiration in the notion of indicators, assessing the biodiversity of certain communities both in terms of taxonomic diversity and their species interactions. The focus would be on a highly diverse group of organisms in a given environment with a range of diverse but representative ecological interactions. Biodiversity assessment on such a group would serve as scaffold to anchor successive complementary studies above and below that particular interaction level, aiming with time at a multitrophic level description of the whole system. In this respect, for terrestrial ecosystems, ubiquitous herbivore insects constitute an excellent focal group to launch large-scale enquiry on the biodiversity and interactions of biomes (Stork and Habel 2014). In our opinion, phytophagous beetles, and leaf beetles in particular, represent a study system with important advantages. Their taxonomic diversity and that

of their food-plants can be staggering in any given tropical environment (Erwin 1982; Wagner 1999; Novotny and Miller 2014), in general they portray a tight ecological relationship with plants in all life cycle stages, and have high endemism rates, both factors generating a perception of strong relationship with the environment. All in all, by focusing on the inventory and interactions of leaf beetles, it is possible to design research simultaneously on two highly diverse components (=indicators) of biota from most tropical ecosystems—insects and plants—, as well as on one of the predominant ecological interactions, herbivory (Price 2002).

Over the past few years (since 2007) we have thus developed on the notion that we can significantly contribute to an enhancement of biodiversity studies by targeting the fast characterisation of complex leaf beetle (or other herbivore insects) communities in the tropics as well as their ecological associations by using a combination of DNA-barcodes, tree-based species delimitation and forensic characterisation of food plants, with a robust and automatable analytical set-up. As a general proposition, we advocate that, when attempting large-scale biodiversity studies, where both delimitation and identification of species represent a challenge, the most efficient approach involves the use of DNA sequence data (only one or few ‘barcodes’) and phylogenetic approaches. Thus, our general workflow for large-scale biodiversity assessment of tropical leaf beetle communities includes four distinctive stages: (1) indiscriminate sampling of chrysomelid beetles in a particular environment or region; (2) non-destructive DNA extractions and specimen preparation for future reference; (3) DNA sequencing of at least one beetle mtDNA marker (typically *cox1*) and at least one putative diet marker (either *trnL* or *psbA-trnH*); and (4) phylogenetic inference for beetle species delimitation and host-plant identification.

## 2.2. Species delimitation and enhanced species discovery

We mentioned above that DNA-enhanced species delimitation has achieved fundamental progress over the past few years in great part thanks to the development of powerful phylogenetic methodologies to deal with gene tree incongruence as well as conceptual advancement on how to integrate taxonomically relevant data. However, these procedures are time and resource consuming, benefiting from the analysis of multiple genes and generally from a good taxonomic knowledge of the group of interest. These tree-based procedures find a good use in systematic research but are impractical for large-scale, rapid biodiversity assessment. Instead, our methods of choice, with a good trade-off between economy and speed of analysis (including data acquisition) and robustness and accuracy of results are the Generalized Mixed Yule-Coalescent model (GMYC; Pons et al. 2006; Fujisawa and Barraclough 2013) and the Bayesian implementation of the Poisson tree processes model for species delimitation (bPTP; Zhang et al. 2013). These are tree-based methods that do not require previous knowledge of species boundaries, making them suitable for the analysis of groups with poor taxonomy, and are specifically designed to work with single locus data (e.g., a DNA-barcode).



For example, GMYC tests changes in branching rates at the species boundary on an ultrametric tree based on the optimisation of a likelihood function with predictions for branching patterns both in speciation and population neutral coalescent processes. In practice, the algorithm scans two types of information on gene trees—waiting times between successive branching events and number of lineages within each interval—to optimise a single or multiple thresholds defining species branches on the tree subtending one or more populations evolving under neutral coalescent diversification processes. bPTP in turn relaxes the need for an ultrametric tree and infers species boundaries based on the so-called Poisson tree processes model (Zhang et al. 2013). Focusing on a single standard DNA-barcode lowers the cost and increases the speed and robustness of data acquisition, and both algorithms are fast and accessible thanks to functions of the R package ‘splits’ (SPecies LImits by Threshold Statistics; Ezard et al. 2013) in the case of GMYC, and a fully functional web server (<http://species.h-its.org/ptp/>) in the case of bPTP, both desirable characteristics for rapid biodiversity assessment.

The suitability of this approach to investigate well-known leaf beetle communities in temperate regions has been shown recently (Baselga et al. 2015). In addition, we are successfully applying it to several projects studying leaf beetle biodiversity at large in different tropical systems for which there is a deficient taxonomic knowledge on the composition of their respective leaf beetle communities. One such study focuses on the diversity of Eumolpinae in New Caledonia, a group that recent taxonomic work has exposed as highly diverse without a precise estimate of the expected total diversity (Gómez-Zurita 2011; Papadopoulou et al. 2013). In other studies we investigate the communities of leaf beetles in dry tropical forests of Nicaragua and Vietnam with a common aim of evaluating biodiversity parameters that can be eventually used for conservation initiatives targeting this highly threatened tropical biome (Janzen 1988; Miles et al. 2006). In these studies, we sampled hundreds of leaf beetle specimens which were individually characterised for one mtDNA standard locus, an 830 bp fragment of the 3'-end of the first subunit of the cytochrome *c* oxidase (*cox1*), and an additional mtDNA locus in the case of New Caledonian Eumolpinae (a 515 bp fragment of the small rRNA subunit, *rrnS*). In every case, the individuals characterised from a genetic viewpoint were preserved and mounted dry, with their genitalia dissected. Vouchering specimens from such large-scale biodiversity studies is essential for fulfilling the inventorying angle of biodiversity assessment, particularly when the lack of readily available taxonomic expertise or the weak taxonomic knowledge of the focal group, hampers the immediate naming of species. The amount of new species for Science in understudied tropical faunas can be high, and subsequent in-depth taxonomic work to name species usually reveals undescribed diversity. As will be seen below, the non-destructive treatment of samples is crucial to allow for species descriptions and instantly provides with standard type material (besides the DNA sequences used to speed up their discovery). Preparation of our processed specimens has yet another short-term practical advantage, which is allowing for a fast complementary assessment of species diversity based on the concept of morphospecies, i.e. groups of individuals that look alike. A comparison between the two pragmatic strategies for rapid species

assessment, DNA-based GMYC-groups *versus* morphospecies, can assist in the evaluation of performance of the first, objective method (Papadopoulou et al. 2013), as well as the discovery of new species, while drawing attention to interesting biological characteristics of the system, particularly if sample metadata is taken into account (e.g., geography, biome, host-plant information, etc.).

The systematic implementation of GMYC species delimitation to each of our datasets produced consistently species counts compatible with estimates based on morphospecies assessment (Table 1), and disagreements revealed in general a better performance of the molecular tree-based strategy. Essentially identical results have been shown and the same perception championed by Tänzler et al. (2011) based on their rapid-biodiversity assessment exercise centred on a single hyper-diverse weevil genus in New Guinea, *Trigonopterus*. Additionally, these authors formally explored a very interesting aspect of rapid species assessment that we also experienced from a pragmatic viewpoint, adding to the value of molecular approaches: DNA-based species delimitation outperforms sorting skills by trained, but non-expert parataxonomists. In our experience, there are always a few cases of morphospecies misplacements that benefit from reassessment *a posteriori* using phylogenetic information. These misplacements are not necessarily the result of real identification difficulties, but could be simply owing to visual memory limitations, when dealing with hundreds, perhaps thousands of specimens belonging to dozens or hundreds of species, in the context of massive sampling in tropical settings. Of course, DNA-based approaches have shown their strength in revealing hidden, cryptic diversity, externally invisible to expert eyes, let alone to rapid sorting for accelerated biodiversity inventories (e.g., *Astraptes*, Prado et al. 2011; Staphylinidae, Thormann et al. 2011). However, there is an additional important advantage of using DNA for species delimitation, somehow tackling the opposite scenario offered by cryptic diversity. This is the opportunity to sort accurately all life-stages (e.g., Ahrens et al. 2007), species with colour polymorphism (e.g., Rugman-Jones et al. 2013) or sexually dimorphic species (e.g., Smith and Brown 2008), i.e. situations that are challenging for morphospecies-based assessment of diversity, while they are rather common in insects, in particular in certain groups such as butterflies and many beetles, including the Chrysomelidae. Our research on tropical leaf beetle communities has

**Table 1.** Sampling and sequencing effort, and DNA-based species diversity estimates in three large-scale leaf beetle biodiversity studies in the tropics.

Study	N	Geographic scope	Longest transect	Taxonomic rank	DNA-barcode	GMYC species
New Caledonia	840	Grande Terre	400 km	Eumolpinae	<i>cox1</i> , <i>rrnS</i>	107 [94-121] <sup>a</sup>
Nicaragua	1270	Pacific and northern provinces	250 km	Cassidinae, Eumolpinae, Galerucinae,	<i>cox1</i>	336 [333-347]
Vietnam	494	Núi Chúa Natl. Pk.	5 km	Chrysomelidae	<i>cox1</i>	161 [156-165] <sup>b</sup>

<sup>a</sup>Averaged data from Papadopoulou et al. (2013).  
<sup>b</sup>Taken and averaged from Nguyen and Gómez-Zurita (in prep.)

provided with examples for each of these advantages, matching larvae and adults of the chrysomeline *Plagiodera septemvittata* Stål in Vietnam (Nguyen and Gómez-Zurita, in prep.) or the cassidines *Coptocycla leprosa* (Boheman), *Omocerus caeruleopunctatus* (Boheman) and *Parorectis rugosa* (Boheman) in Nicaragua (Papadopoulou et al. 2015), the very distinctive males and females of several eumolpine species of *Taophila* Heller in New Caledonia (Papadopoulou et al. 2013; Gómez-Zurita and Cardoso 2014), and the highly polymorphic galerucine *Cerotoma atrofasciata* Jacoby in Nicaragua.

Once there is a sound estimate of species numbers resulting from a sampling effort of known intensity, it is possible to investigate how representative the measure of biodiversity is of the total expected diversity. For example, we used a strategy based on rarefaction curves representing accumulation of objectively delimited species across sampling events for New Caledonian Eumolpinae to extrapolate the expected total species richness in the studied environments. From our empirical demonstration of slightly over one hundred species in our ensemble sample, we could analytically propose an expected total number of eumolpine species in New Caledonia between 148 and 210, depending on input data and species richness estimator of choice (Papadopoulou et al. 2013). Preliminary data for three Chrysomelidae subfamilies sampled in Nicaraguan dry forests (Eumolpinae, Cassidinae s.l. and Galerucinae s.l.) or the whole Chrysomelidae community in a National Park in southern Vietnam, both analysed using a similar accumulation-curve approaches as in New Caledonia, reveal that our samples may represent between 53–69% of the total leaf beetle diversity in the studied biomes. Thus, a continued sampling effort should recognise in the order of 500–600 Chrysomelidae species in the abovementioned subfamilies in the dry Pacific side of Nicaragua, or the same number of chrysomelids in a 10 sq. km. forest patch across a slight elevation gradient in southern Vietnam.

The experience gained from this type of studies shows that the main limiting factor for robust diversity assessment is obtaining sampling densities representative of the studied environment always, i.e. fieldwork. Once samples are available, laboratory methods can be optimised in weeks or few months, depending on the number of samples used and smoothness of PCR protocols, and a similar or slightly longer time for standardised analytical procedures.

### 2.3. Forensic methods for the analysis of species interactions

We stressed already that there is one quantitative advantage of molecular characters to aid biodiversity assessment: speeding up the rate of species delimitation and also diagnosis. Additionally, these characters have at the same time the potential to contribute an extremely important qualitative advantage: the possibility to investigate complex systems and processed samples, which is the door to community ecology and the study of food-webs. In 2009, simultaneously with the studies of Valentini et al. (2009) and Soininen et al. (2009) and the earlier approach of Matheson et al. (2008), we pioneered the investigation of animal-plant interactions using DNA (Jurado-Rivera et al.

2009). In our approach, conversely to the mainstream DNA-barcoding stance of these contemporaneous and other subsequent studies, one of our main motivations was to extract taxonomically relevant information from processed food in the face of an incomplete reference database, by exploiting molecular phylogenies as the most rigorous and powerful tool for taxonomic assessment.

In most studies that target trophic associations, DNA extraction is directed to the most obvious sources for food DNA, including gut contents and faeces. In our case, and in great part motivated by the special characteristics of our study organism, the starting material is always the whole leaf beetle specimen, generally small enough to fit the tubes used for the DNA extraction procedure. The main idea is that when we obtain DNA from the whole specimen, we indeed mostly retrieve nucleic acids from the beetle species, useful for its genetic characterisation. However, with host DNA, we obtain simultaneously a significant proportion of DNA from organisms onto and into the beetle, therefore representing the ecological interactions it sustains, including DNA from all of its symbionts, endosymbionts, phoretics, commensals, parasites, hyperparasites and, of course, food remains. We refer to this condition as the *ecology inside a vial*. In recent years, we have been particularly interested in the analysis of the host trophic ecology, but the same samples are amenable to studies of different trophic levels (see Montagna et al. 2015, for a pioneering study on leaf beetle microbiomes, for instance).

PCR-based molecular characterisation of a predator's food can be challenging, particularly in the case of carnivorous animals, when their food can belong to a closely related taxon, requiring a selective procedure to distinguish (and avoid) template DNA from the host. In a DNA metabarcoding framework, this can be achieved by using primers specifically designed to target a specific taxonomic group of potential diets (e.g., Riaz et al. 2011). One such example is the use of insect-specific mtDNA PCR primers to identify insects preyed by spiders, which takes advantage of the high mtDNA divergence between these two taxonomic Classes, allowing for selective PCR (e.g., Northam et al. 2012; Sint et al. 2015). Yet, even large taxonomic gaps could result in non-trivial technicalities hampering the design of suitable primers. However, the same type of analysis to investigate the diet of a herbivore is much simpler methodologically, since Nature provides already with the best possible tool: plastid DNA (cpDNA), exclusive of plants, and together with ITS sequences, the marker of choice for DNA-based plant identification, as well as for plant DNA-barcoding (Kress et al. 2005). Botanical molecular systematic research has provided through the years with robust universal primers targeting a variety of cpDNA loci to assist plant species diagnosis. Among these, two loci in particular have been selected as the standard for plant DNA-barcoding, the tandem *rbcL* and *matK* (CBOL Plant Working Group 2009). These and other loci are generally easy to amplify with specific and reliable universal primers which are not interfered, by definition, by animal DNA; they produce PCR fragments of suitable size for easy amplification and sequencing; and their continued use by botanists determines a high taxonomic representation in nucleotide sequence databases, which makes them suitable for identification purposes. While



DNA-barcoding has favoured the use of length-invariant, protein coding loci, in our implementation for herbivore diet inferences, we have opted instead for length-variable cpDNA intergenic spacers, specifically the so-called *trnL* intron and most recently the *psbA-trnH* spacer (Jurado-Rivera et al. 2009; Gómez-Zurita and Cardoso 2014; Papadopoulou et al. 2015; De la Cadena et al., 2016). Sequence length differences are a nuisance for similarity assessment and genetic distance estimation and thus impair reliability of fast algorithms for taxonomic assignment. Yet, in our opinion, and specifically from a phylogenetic perspective, sequence length differences can be efficiently treated with current multiple sequence alignment algorithms, and provide with two main advantages: (1) they become an additional source of useful variation to increase the diagnostic value of these markers, and (2) size differences usually enable resolving homologous PCR products from different species by means of agarose gel electrophoresis, allowing to skip expensive and time-consuming cloning steps when studying the diet of leaf beetle individuals that fed upon two or more plant species.

We showed that this methodology is efficient and highly informative based on our extensive study of diets of Australian Chrysomelinae (Jurado-Rivera et al. 2009). In that study, we used *trnL* sequences obtained from whole specimen DNA extractions to infer the diet of 76 species in 24 genera of Chrysomelinae based on individual phylogenetic analyses carefully including all closely related homologous sequences available in GenBank at the time. In this proof-of-principle study, we were able to infer the correct host plant family in every case (for many species we had known host records), although resolution dropped at lower taxonomic levels (83% at tribal, and 51% at generic levels). Robust phylogenetic analyses provided a sound identification shortcut relying on information available in public sequence databases, and despite lower accuracy at infrafamilial taxonomic levels, we could refine our inferences, sometimes down to the species level, thanks to detailed floristic catalogues for the areas where the beetles had been collected. Yet public database incompleteness is a severe problem and inference power greatly benefits from availability of a local reference database for meaningful comparisons (e.g., García-Robledo et al. 2013). Indeed, since 2008 we have been working on setting a standard for this type of analyses whereby the analysis of the leaf beetle community goes hand-in-hand with a systematic compilation of angiosperm sequence data from the biome of interest to provide with a sound reference library for DNA-based inference of ecological associations. In the particular case of Nicaragua, we have sampled, vouchered, sequenced and made available to the scientific community *psbA-trnH* sequence data for some 450 plant species, nearly half of the plant diversity present in the Nicaraguan dry biomes, in an ongoing effort to enhance DNA-based species identification that we can use to characterise these valuable environments (Papadopoulou et al. 2015).

These approaches are becoming standard in many studies of tropical biodiversity, including studies on leaf beetles (Table 2) and other groups of phytophagous beetles, mainly weevils (Pinzón-Navarro et al. 2010; Kitson et al. 2013). But precisely in the context of large-scale and rapid biodiversity assessment, the generalisation of this type of studies is generating a new challenge. In our specific study of dry tropical forest

structure and interactions in Nicaragua, we have analysed some 840 individual leaf beetle specimens, which yielded nearly 1100 sequences of putative diets. Such a large amount of data is not anymore amenable to individualised tree-based inferences, and two alternatives stand out to scale-up accelerated biodiversity assessment: either giving up trees and using fast BLAST-based approaches or, alternatively, automating the inference process. Given our concerns about the unavoidable problem of incomplete reference databases, especially when working at a regional scale or above, we have opted for the latter. Automated taxonomic identification from multiple sequences can be efficiently tackled by splitting the data into phylogenetically robust datasets together with taxonomically-tagged homologs from GenBank and/or a local reference database of known taxonomy. Making this procedure fully automated meets two main challenges: one is extracting this meaningful subset of homologs and their taxonomically relevant information, and the other is parsing phylogenetic trees for taxonomic information. We have developed a dynamic procedure that solves these problems in efficient ways to iteratively generate tree-based taxonomic identifications from large collections of unidentified DNA-barcoding data, which we called BAGpipe ('Pipeline for Biodiversity Assessment using Genbank data'; Papadopoulou et al. 2015). Starting from a collection of sequence data of the selected genetic marker, the procedure uses a combination of local and global similarity searches to pick up all similar and putatively homologous sequences available in the latest Genbank release, recording their taxon ID and associated taxonomic hierarchy. At the same time, sequences are reoriented if needed, their ends trimmed to the length of the marker of choice, and redundant sequence data (i.e., population data) removed. These ensemble data constitutes the basis for subsequent phylogenetic matrix assemblage and phylogenetic inference, the so-called *reference database*. Robust phylogenetic inference is achieved for a certain level of sequence divergence where positional homology assessment is not compromised and homoplasy due to saturation is low (Goldman 1998; Yang 1998). Thus, we solved the problem of data partition for meaningful phylogenetic inferences by first splitting the unidentified *query sequences* in groups of similarity below a custom divergence threshold, each one used in turn to extract similar sequences from the reference database based on the same criterion. Query sequences and taxonomically identified reference database sequences within a predefined divergence threshold are submitted to multiple sequence alignment and maximum likelihood tree inference (and node support assessment). Automatically drawing taxonomic conclusions from trees was a challenge that we met exploiting the taxonomic hierarchy attached to Genbank data (inspired by Hunt and Vogler 2008; Chesters and Vogler 2013). The obtained unrooted trees are secondarily polarised and the most inclusive supported clades including unidentified query sequences are recognised, parsing the common taxonomy from reference sequences (i.e. Genbank taxon IDs and their hierarchy). This taxonomic inference, at the lowest taxonomic level allowed by the reference, is finally linked to the unidentified query sequence(s) using both strict and liberal criteria. In this context, it becomes obvious that coverage and reliability of available barcode reference libraries are critical for a meaningful use of this approach (Jinbo et al. 2011). A tool like BAGpipe

**Table 2.** Molecular analyses of insect-plant associations for tropical Chrysomelidae.

Leaf beetle	Source	cpDNA marker	Host-plant	Reference
<i>Alagoasa decemguttata</i>	Nicaragua	<i>psbA-trnH</i>	Verbenaceae, Bignoniaceae	De la Cadena et al. (2016)
<i>Anadimonia</i> sp.	Borneo	<i>rbcL</i>	Lauraceae, Dipterocarpaceae	Kishimoto-Yamada et al. (2013)
<i>Arsipoda geographica</i>	New Caledonia	<i>trnL</i>	<i>Ardisia</i> (Myrsinaceae)	Gómez-Zurita et al. (2010)
<i>A. isola</i>	New Caledonia	<i>trnL</i>	Ericaceae	Gómez-Zurita et al. (2010)
<i>Blepharida suturalis</i>	Nicaragua	<i>psbA-trnH</i>	Burseraceae, Boraginaceae	De la Cadena et al. (2016)
<i>Brachycoryna pumila</i>	Nicaragua	<i>psbA-trnH</i>	<i>Sida</i> and <i>Triumfetta</i> (Malvaceae), <i>Chiococca</i> (Rubiaceae)	Papadopoulou et al. (2015)
<i>Calligrapha thermalis</i>	Mexico	<i>psbA-trnH</i>	<i>Perymenium</i> (Asteraceae)	Montelongo and Gómez-Zurita (2013)
<i>Cephaloleia</i> spp.	Costa Rica	ITS2, <i>rbcL</i>	Heliconiaceae, Zingiberaceae, Costaceae, Marantaceae, Cannaceae	García-Robledo et al. (2013)
<i>Chelobasis bicolor</i>	Costa Rica	<i>rbcL</i>	<i>Heliconia</i> (Heliconiaceae)	García-Robledo et al. (2013)
<i>C. perplexa</i>	Costa Rica	<i>rbcL</i>	<i>Heliconia</i> (Heliconiaceae)	García-Robledo et al. (2013)
<i>Dematochroma cancellata</i>	New Caledonia	<i>psbA-trnH</i>	Primulaceae, Lamiaceae, Millettieae (Fabaceae), Cunoniaceae, <i>Syzygium</i> (Myrtaceae), Sapindoideae (Sapindaceae)	Gómez-Zurita and Cardoso (2014)
<i>Glenidion</i> sp.	Nicaragua	<i>psbA-trnH</i>	Burseraceae, Fabaceae, Lantaneae (Verbenaceae)	De la Cadena et al. (2016)
<i>Heterispa vinula</i>	Nicaragua	<i>psbA-trnH</i>	Malvaceae, Cucurbitaceae, Annonaceae, Poaceae, Boraginaceae	Papadopoulou et al. (2015)
<i>Hyphaenia</i> sp.	Borneo	<i>rbcL</i>	polyphagous (7 plant families)	Kishimoto-Yamada et al. (2013)
<i>Liroetiella antennata</i>	Borneo	<i>rbcL</i>	Acanthaceae, Fabaceae, Fagaceae, Moraceae	Kishimoto-Yamada et al. (2013)
<i>Monolepta</i> spp.	Borneo	<i>rbcL</i>	polyphagous	Kishimoto-Yamada et al. (2013)
<i>Omophoita octomaculata</i>	Nicaragua	<i>psbA-trnH</i>	<i>Stachytarpheta</i> (Verbenaceae), Lamiaceae	De la Cadena et al. (2016)
<i>Parorectis rugosa</i>	Nicaragua	<i>psbA-trnH</i>	<i>Physalis</i> and <i>Solanum</i> (Solanaceae), Lamiaceae, Cucurbitaceae, Scrophulariaceae, Fabaceae	Papadopoulou et al. (2015)
<i>Physonota alutacea</i>	Nicaragua	<i>psbA-trnH</i>	<i>Cordia</i> (Boraginaceae), Fabaceae	Papadopoulou et al. (2015); De la Cadena et al. (2016)
<i>Platymela cephalotes</i>	Australia	<i>trnL</i>	<i>Acacia</i> (Fabaceae)	Jurado-Rivera et al. (2009)

<i>Syphrea</i> sp.	Nicaragua	<i>psbA-trnH</i>	<i>Acalypha</i> (Euphorbiaceae)	De la Cadena et al. (2016)
<i>Taophila</i> ( <i>Jolivetiana</i> ) <i>mantillerii</i>	New Caledonia	<i>psbA-trnH</i>	Polypodiopsida	Gómez-Zurita and Cardoso (2014)
<i>Taophila</i> ( <i>Lapita</i> ) spp.	New Caledonia	<i>psbA-trnH</i>	Cyatheales, Fabaceae, <i>Syzygium</i> (Myrtaceae), Rauvolfioideae (Apocynaceae), Oxalidales, Sterculioideae (Malvaceae)	Gómez-Zurita and Cardoso (2014)
<i>Taophila</i> s. str. spp.	New Caledonia	<i>psbA-trnH</i>	Polypodiopsida, Primulaceae, Millettieae (Fabaceae)	Gómez-Zurita and Cardoso (2014)
<i>Theopea</i> sp.	Borneo	<i>rbcL</i>	polyphagous (10 plant families)	Kishimoto-Yamada et al. (2013)
<i>Walterianella venustula</i>	Nicaragua	<i>psbA-trnH</i>	Lamiaceae, <i>Buddleja</i> (Scrophulariaceae)	De la Cadena et al. (2016)

(<http://www.ibe.upf-csic.es/SOFT/Softwareanddata.html>) makes it possible to boost large-scale biodiversity assessment both in its inventorying angle, but also in the study of interactions if applied to the identification of ecology-in-a-vial associations or metabarcoding studies.

2.4. Simultaneous progress in inventory and interactions

From our previous account, it should be clear already that the use of DNA has the potential to enhance simultaneously the study of both species inventories and species interactions, by using a limited number of standard laboratory and analytical techniques. In molecular systematics research, it is routine to use the PCR technique and suitable sets of primers to amplify more than one molecular marker from each sample. These data combined inform on the organisation of diversity and can potentially hint at specific evolutionary processes that shaped this diversity. Based on this common practice, we have easily incorporated to the lab routine the characterisation of a plant cpDNA marker from leaf beetle DNA extractions, in addition to our standard beetle markers. As a result, we systematically add a new ecological dimension to the description of diversity. We described several new tropical leaf beetle species interpreting DNA differences with other known beetle taxa, providing also with a DNA-based diagnosis of plant species for putative diet sequences. These include a southern Nearctic Chrysomelinae, the Mexican *Calligrapha thermalis* Gómez-Zurita associated to the composite *Perymenium mendezii* (Montelongo and Gómez-Zurita 2013), two species of New Caledonian Alticinae in the genus *Arsipoda* with one of them associated to Myrsinaceae (Gómez-Zurita et al. 2010), and two species of the New Caledonian endemic genus *Taophila* (Eumolpinae) together with an assessment of their dietary breadth (Gómez-Zurita and Cardoso 2014).

The above examples do not fall of course in the category of large-scale biodiversity assessment, although at least in the particular case of the study on the genus *Taophila*,



it is a direct consequence, a refinement of findings derived from the wider biodiversity scope facilitated by this methodological approach (Papadopoulou et al. 2013). Nevertheless, each of these studies contributes individually to our understanding of tropical biodiversity and, if this strategy became the standard for systematic research in herbivore beetles, it would represent a fast progress in the complementary analysis of species and interactions. As seen, scaling-up this strategy for community analyses is feasible. Yet, we strongly believe that, even if some steps in species delimitation and identification are facilitated by the use of the described techniques, there will always be a dramatic need of taxonomic expertise to come full circle in any attempt for reliable biodiversity assessment.

### 3. Concluding remarks

As a short summary of our contribution, we can highlight that biodiversity is more than just species lists, and that biodiversity assessment should not neglect the way in which species are inter-connected in the ecosystems. Cataloguing biodiversity at large is certainly challenging, but it is also feasible, and DNA is possibly the key to fast and as comprehensive as possible inventorying of life forms, but also of their interactions. Phylogenies provide a robust approach to species delimitation and, in the absence of a comprehensive reference for comparison, the most robust approach to DNA-based species identification. Finally, the use of DNA as standard for species delimitation and identification makes these processes fully automatable, which is essential for high-throughput biodiversity assessment.

We tried to be constructive and discuss solutions to some of the current challenges in large-scale biodiversity assessment, however some fundamental problems remain and are not exclusively conditioned by technological or conceptual advancement. Rather, societal awareness (which is in great part our responsibility as professionals of biodiversity) and commitment of politicians and funding agencies alone can provide already a quantitative advantage for biodiversity research. As noted before, the emphasis for effective biodiversity research needs to be put again on funding expeditions and environmental sampling, pretty much with the same spirit as in the original voyages of discovery, but with the benefits of technology and trained specialists in different groups. Initiatives of this kind exist, most notably targeting insular systems, e.g. SANTO 2006, targeting the island of Espiritu Santo, the largest in the archipelago of Vanuatu (<http://www.santo2006.org>), or the Mo'orea Biocode Project, on the homonym island in the Tahiti archipelago (<http://mooreabiocode.org>). While these initiatives exceptionally mobilise millions of dollars and hundreds of scientists for comprehensive biological prospection, and are built with the right spirit, they typically yield a very modest global output. The reason is the currently existing bottleneck of available taxonomic expertise for extracting meaningful biodiversity information from these surveys, which remains the most serious challenge for large-scale biodiversity research (Kim and Byrne 2006). This towering limitation impairs not only the rigor-

ous assessment of biodiversity in classical ways, but also our chances to count with a reliable taxonomy attached to public sequence databases, one of the most valuable resources for improving and speeding-up biodiversity assessment. Again, this challenge can be in part solved by restoring the importance and value of taxonomic research and allocating resources to taxonomic training, coupled with commitment and pedagogy for and from taxonomists to reinforce and expand available expertise.

Besides these fundamental limitations, there are still others of technical and conceptual nature which need to be dealt with, such as devising creative and efficient ways to incorporate new technologies for the improvement of large-scale biodiversity assessment. These should include for instance the use of next-generation sequencing technologies and environmental metagenomics, or more specifically in the case of insects the recently developed ‘metagenome skimming’ approach (Andújar et al. 2015; Linard et al. 2015), which promises to transform the standards of DNA-based biodiversity assessment by eliminating the PCR step and associated biases. Additionally, new automated procedures are required to democratise both species delimitation and identification (through reliable publicly available references). Of course, as conveyed in this assay from the start, there must be also a dedicated effort to routinely integrate ideas of inventory and interactions in biodiversity surveys, with an ever larger and more integrative scope. These and many more ideas are in the agenda of biodiversity researchers, as evidenced by many international Biodiversity Initiatives throughout the world and at different scales, of which these with global scope are the paradigm for large-scale biodiversity assessment, e.g. Global Ocean Biodiversity Initiative (<http://www.gobi.org>), Center for Tropical Forest Science (<http://www.ctfs.edu>), or the Global Genome Biodiversity Network (<http://data.ggbn.org/index.php>), among others. These initiatives address some of their objectives by bringing genomics, taxonomy and ecology together through the combination of strategic sampling and massive sequencing technologies, when possible.

## **Acknowledgements**

The ‘Fundación BBVA’ (Spain) has funded the bulk of this work thanks to their support for our large-scale biodiversity assessment initiative in Nicaraguan tropical dry forests (project BIOCON08-045, IP: JGZ). Our work in Nicaragua has benefited from a postdoctoral ‘Juan de la Cierva’ contract (Spanish Ministry of Science and Innovation, MICINN) to AP, and an AECID predoctoral studentship (Spanish Ministry of Foreign Affairs and Cooperation) and a SENESCYT scholarship (Secretariat of High Education, Science, Technology and Innovation, Ecuador) to GDC. The National Geographic Society supported most of our research in New Caledonia (project 8380-07, IP: JGZ) with help from a travel grant awarded by the Percy Sladen Memorial Fund of the Linnean Society of London to JGZ. The Spanish High Research Council (CSIC), in the framework of a cooperation agreement with the Vietnamese Academy of Sciences, supports our work in dry tropical forests of southern Vietnam

(IP: JGZ) as well as a predoctoral studentship to DTN. Several EU Synthesys research stays (GB-TAF-1840, SE-TAF-1893, DE-TAF-4348) and a Mayr Travel Grant (Harvard University) as well as project CGL2008-00007/BOS (MICINN, IP: JGZ) have contributed to the discovery of a new tropical species of *Calligrapha*, and the latter also framed the predoctoral studentship to TM.

## References

- Ahrens D, Monaghan MT, Vogler AP (2007) DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Molecular Phylogenetics and Evolution* 44: 436–449. doi: 10.1016/j.ympev.2007.02.024
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389–3402. doi: 10.1093/nar/25.17.3389
- Andújar C, Arribas P, Ruíz C, Serrano J, Gómez-Zurita J (2014) Integration of conflict into Integrative Taxonomy: fitting hybridization in the delimitation of species in *Mesocarabus* (Coleoptera: Carabidae). *Molecular Ecology* 23: 4344–4361. doi: 10.1111/mec.12793
- Andújar C, Arribas P, Ruzicka F, Crampton-Platt A, Timmermans MJ, Vogler AP (2015) Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology* 24: 3603–3617. doi: 10.1111/mec.13195
- Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, Mersey B, Ferrer EA (2011) Has the Earth's sixth mass extinction already arrived? *Nature* 471: 51–57. doi: 10.1038/nature09678
- Baselga A, Gómez-Rodríguez C, Vogler AP (2015) Multi-hierarchical macroecology at species and genetic levels to discern neutral and non-neutral processes. *Global Ecology and Biogeography* 24: 873–882. doi: 10.1111/geb.12322
- Basset Y, Missa O, Alonso A, Miller SE, Curletti G, De Meyer M, Eardley CD, Mansell MW, Novotny V, Wagner T (1998) Faunal turnover of arthropod assemblages along a wide gradient of disturbance in Gabon. *African Entomology* 16: 47–59. doi: 10.4001/1021-3589-16.1.47
- Berger SA, Krompass D, Stamatakis A (2011) Performance, accuracy, and Web server for evolutionary placement of short sequence reads under Maximum Likelihood. *Systematic Biology* 60: 291–302. doi: 10.1093/sysbio/syr010
- Caro TM, O'Doherty G (1999) On the use of surrogate species in conservation biology. *Conservation Biology* 13: 805–814. doi: 10.1046/j.1523-1739.1999.98338.x
- CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences USA* 106: 12794–12797. doi: 10.1073/pnas.0905845106
- Chesters D, Vogler AP (2013) Resolving ambiguity of species limits and concatenation in multilocus sequence data for the construction of phylogenetic supermatrices. *Systematic Biology* 62: 456–466. doi: 10.1093/sysbio/syt011
- Dayrat B (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85: 407–415. doi: 10.1111/j.1095-8312.2005.00503.x

- De la Cadena G, Papadopoulou A, Maes J-M, Gómez-Zurita J (2016) Evaluation of bias on the assessment of diet breadth of herbivorous insects using molecular methods. *Insect Science*. doi: 10.1111/1744-7917.12303
- Dirzo R, Raven PH (2003) Global state of biodiversity and loss. *Annual Review of Environment and Resources* 28: 137–167. doi: 10.1146/annurev.energy.28.050302.105532
- Dowle EJ, Morgan-Richards M, Trewick SA (2013) Molecular evolution and the latitudinal biodiversity gradient. *Heredity* 110: 501–510. doi: 10.1038/hdy.2013.4
- Dunn RR (2005) Modern insect extinctions, the neglected majority. *Conservation Biology* 19: 1030–1036. doi: 10.1111/j.1523-1739.2005.00078.x
- Ekrem T, Willassen E, Stur E (2007) A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution* 43: 530–542. doi: 10.1016/j.ympev.2006.11.021
- Erwin TL (1982) Tropical forests: their richness in Coleoptera and other arthropod species. *Coleopterists Bulletin* 36: 74–75.
- Ezard T, Fujisawa T, Barraclough TG (2013) splits: SPecies' LLimits by Threshold Statistics. R package version 1.0. <http://R-Forge.R-project.org/projects/splits>
- Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology* 62: 707–724. doi: 10.1093/sysbio/syt033
- Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution* 27: 480–488. doi: 10.1016/j.tree.2012.04.012
- García-Robledo C, Erickson DL, Staines CL, Erwin TL, Kress WJ (2013) Tropical plant-herbivore networks: reconstructing species interactions using DNA barcodes. *PLoS ONE* 8: e52967. doi: 10.1371/journal.pone.0052967
- Goldman N (1998) Phylogenetic information and experimental design in molecular systematics. *Proceedings of the Royal Society of London B* 265: 1779–1786. doi: 10.1098/rspb.1998.0502
- Gómez-Zurita J (2011) Revision of New Caledonian species of Eumolpinae described by K. M. Heller (Coleoptera: Chrysomelidae). *Zootaxa* 3060: 31–46.
- Gómez-Zurita J, Cardoso A (2014) Systematics of the New Caledonian endemic genus *Taophila* Heller (Coleoptera: Chrysomelidae, Eumolpinae) combining morphological, molecular and ecological data, with description of two new species. *Systematic Entomology* 39: 111–126. doi: 10.1111/syen.12038
- Gómez-Zurita J, Cardoso A, Jurado-Rivera JA, Jolivet P, Cazères S, Mille C (2010) Discovery of a new species of New Caledonian *Arsipoda* Erichson, 1842 (Coleoptera: Chrysomelidae) and insights on their ecology and evolution using DNA markers. *Journal of Natural History* 44: 2557–2579. doi: 10.1080/00222933.2010.499575
- Hebert PD, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B* 270: 313–321. doi: 10.1098/rspb.2002.2218
- Hunt T, Vogler AP (2008) A protocol for large-scale rRNA sequence analysis: towards a detailed phylogeny of Coleoptera. *Molecular Phylogenetics and Evolution* 47: 289–301. doi: 10.1016/j.ympev.2007.11.029



- Janzen DH (1973) Comments on host-specificity of tropical herbivores and its relevance to species richness. In: Heywood VH (Ed.) *Taxonomy and ecology*. Systematic Association Special Vol. 5. Academic Press, London, 201–211.
- Janzen DH (1988) Tropical dry forests. The most endangered major tropical ecosystem. In: Wilson EO (Ed.) *Biodiversity*. National Academy Press, Washington, DC, 130–137.
- Jinbo U, Kato T, Ito M (2011) Current progress in DNA barcoding and future implications for entomology. *Entomological Science* 14: 107–124. doi: 10.1111/j.1479-8298.2011.00449.x
- Jurado-Rivera J, Vogler AP, Reid CAM, Petitpierre E, Gómez-Zurita J (2009) DNA barcoding insect-host plant associations. *Proceedings of the Royal Society of London B* 276: 639–648. doi: 10.1098/rspb.2008.1264
- Kim KC, Byrne LB (2006) Biodiversity loss and the taxonomic bottleneck: emerging biodiversity science. *Ecological Research* 21: 794–810. doi: 10.1007/s11284-006-0035-7
- Kishimoto-Yamada K, Kamiya K, Meleng P, Diway B, Kaliang H, Chong L, Itioka Sakai S, Ito M (2013) Wide host ranges of herbivorous beetles? Insights from DNA Bar Coding. *PLoS ONE* 8: e74426. doi: 10.1371/journal.pone.0074426
- Kitson JJN, Warren BH, Vincent Florens FB, Baider C, Strasberg D, Emerson BC (2013) Molecular characterization of trophic ecology within an island radiation of insect herbivores (Curculionidae: Entiminae: *Cratopus*). *Molecular Ecology* 22: 5441–5455. doi: 10.1111/mec.12477
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences USA* 102: 8369–8374. doi: 10.1073/pnas.0503123102
- Linard B, Crampton-Platt A, Timmermans MJTN, Vogler AP (2015) Metagenome skimming of insect specimen pools: potential for comparative genomics. *Genome Biology and Evolution* 7: 1474–1489. doi: 10.1093/gbe/evv086
- Lipscomb D, Platnick N, Wheeler Q (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology and Evolution* 18: 65–68. doi: 10.1016/S0169-5347(02)00060-5
- Matheson CD, Muller GC, Junnila A, Vernon K, Hausmann A, Miller MA, Greenblatt C, Schlein Y (2008) A PCR method for detection of plant meals from the guts of insects. *Organisms, Diversity and Evolution* 7: 294–303. doi: 10.1016/j.ode.2006.09.002
- McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME Journal* 6: 610–618. doi: 10.1038/ismej.2011.139
- Meier R, Shiyang K, Vaidya G, Ng PK (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* 55: 715–728. doi: 10.1080/10635150600969864
- Meyer C, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* 3: e422. doi: 10.1371/journal.pbio.0030422
- Miles L, Newton AC, DeFries RS, Ravilious C, May I, Blyth S, Kapos V, Gordon JE (2006) A global overview of the conservation status of tropical dry forests. *Journal of Biogeography* 33: 491–505. doi: 10.1111/j.1365-2699.2005.01424.x

- Montagna M, Gómez-Zurita J, Giorgi A, Epis S, Lozzia G, Bandi C (2015) Metamicrobiomics in herbivore beetles of the genus *Cryptocephalus* (Chrysomelidae): toward the understanding of ecological determinants in insect symbiosis. *Insect Science* 22: 340–352.
- Montelongo T, Gómez-Zurita J (2013) Morphological and molecular characterization of a new nearctic species of *Calligrapha* (Coleoptera: Chrysomelidae, Chrysomelinae) from Central Mexico. *Proceedings of the Entomological Society of Washington* 115: 369–391. doi: 10.4289/0013-8797.115.4.369
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858. doi: 10.1038/35002501
- Northam WT, Allison LA, Cristol DA (2012) Using group-specific PCR to detect predation of mayflies (Ephemeroptera) by wolf spiders (Lycosidae) at a mercury-contaminated site. *Science of the Total Environment* 416: 225–231. doi: 10.1016/j.scitotenv.2011.11.083
- Noss RF (1990) Indicators for monitoring Biodiversity- A hierarchical approach. *Conservation Biology* 4: 355–364. doi: 10.1111/j.1523-1739.1990.tb00309.x
- Novotny V, Miller SE (2014) Mapping and understanding the diversity of insects in the tropics: past achievements and future directions. *Austral Entomology* 53: 259–267. doi: 10.1111/aen.12111
- Oliver I, Beattie AJ (1996) Designing a cost-effective invertebrate survey: a test of methods for rapid assessment of biodiversity. *Ecological Applications* 6: 594–607. doi: 10.2307/2269394
- Papadopoulou A, Cardoso A, Gómez-Zurita J (2013) Diversity and diversification of Eumolpinae (Coleoptera: Chrysomelidae) in New Caledonia. *Zoological Journal of the Linnean Society* 168: 473–495. doi: 10.1111/zoj.12039
- Papadopoulou A, Chesters D, Coronado I, De la Cadena G, Cardoso A, Reyes JC, Maes J-M, Rueda RM, Gómez-Zurita J (2015) Automated DNA-based plant identification for large-scale biodiversity assessment. *Molecular Ecology Resources* 15: 136–152. doi: 10.1111/1755-0998.12256
- Pianka ER (1966) Latitudinal gradients in species diversity: a review of concepts. *American Naturalist* 100: 33–46. doi: 10.1086/282398
- Pinzón Navarro S, Jurado-Rivera JA, Gómez-Zurita J, Lyal CHC, Vogler AP (2010) DNA profiling of host-herbivore interactions in tropical forests. *Ecological Entomology* 35: 18–32. doi: 10.1111/j.1365-2311.2009.01145.x
- Prado BR, Pozo C, Valdez-Moreno M, Hebert PDN (2011) Beyond the colours: discovering hidden diversity in the Nymphalidae of the Yucatan Peninsula in Mexico through DNA barcoding. *PLoS ONE* 6: e27776. doi: 10.1371/journal.pone.0027776
- Price PW (2002) Species interactions and the evolution of biodiversity. In: Herrera CM, Pellmyr O (Eds) *Plant-Animal Interactions: An Evolutionary Approach*. Blackwell Science, Oxford, 3–25.
- Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic Barcode Gap Discover for primary species delimitation. *Molecular Ecology* 21: 1864–1877. doi: 10.1111/j.1365-294X.2011.05239.x
- Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research* 39: e145. doi: 10.1093/nar/gkr732

- Ross HA, Murugan S, Sibon Li WL (2008) Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology* 57: 216–230. doi: 10.1080/1063-5150802032990
- Rugman-Jones PF, Hoddle CD, Hoddle MS, Stouthamer R (2013) The lesser of two weevils: molecular-genetics of pest palm weevil populations confirm *Rhynchophorus vulneratus* (Panzer 1798) as a valid species distinct from *R. ferrugineus* (Olivier 1790), and reveal the global extent of both. *PLoS ONE* 8: e78379. doi: 10.1371/journal.pone.0078379
- Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH (2010) Integrative Taxonomy: A multisource approach to exploring biodiversity. *Annual Review of Entomology* 55: 421–438. doi: 10.1146/annurev-ento-112408-085432
- Singh JS (2002) The biodiversity crisis: A multifaceted review. *Current Science* 82: 638–647.
- Sint D, Thurner I, Kaufmann R, Traugott M (2015) Sparing spiders: faeces as a non-invasive source of DNA. *Frontiers in Zoology* 12: 3. doi: 10.1186/s12983-015-0096-y
- Smith PT, Brown BV (2008) Utility of DNA sequences for inferring phylogenetic relationships and associating morphologically dissimilar males and females of the bee-killing flies, genus *Melaloncha* (Diptera: Phoridae). *Annals of the Entomological Society of America* 101: 713–721. doi: 10.1603/0013-8746(2008)101[713:UODSFI]2.0.CO;2
- Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebo JH, Ims RA, Yoccoz NG, Taberlet P (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Frontiers in Zoology* 6: 16. doi: 10.1186/1742-9994-6-16
- Stork NE, Habel JC (2014) Can biodiversity hotspots protect more than tropical forest plants and vertebrates? *Journal of Biogeography* 41: 421–428. doi: 10.1111/jbi.12223
- Tänzler R, Sagata K, Surbakti S, Balke M, Riedel A (2012) DNA barcoding for Community Ecology - How to tackle a hyperdiverse, mostly undescribed Melanesian fauna. *PLoS ONE* 7: e28832. doi: 10.1371/journal.pone.0028832
- Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution* 18: 70–74. doi: 10.1016/S0169-5347(02)00041-1
- Thormann B, Raupach MJ, Wagner T, Wägele JW, Peters MK (2011) Testing a short nuclear marker for inferring staphylinid beetle diversity in an African tropical rain forest. *PLoS ONE* 6(3): e18101. doi: 10.1371/journal.pone.0018101
- Timmermans MJ, Dodsworth S, Culverwell C, Bocak L, Ahrens D, Littlewood DT, Pons J, Vogler AP (2010) Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Research* 38: e197. doi: 10.1093/nar/gkq807
- Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Pompanon F, Gielly L, Cruaud C, Nascetti G, Wincker P, Swenson JE, Taberlet P (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach. *Molecular Ecology Resources* 9: 51–60. doi: 10.1111/j.1755-0998.2008.02352.x
- Wagner T (1999) Arboreal chrysomelid community structure and faunal overlap between different types of forests in Central Africa. In: Cox ML (Ed.) *Advances in Chrysomelidae Biology* 1. Backhuys Publishers, Leiden, 247–270.

- Western D (1992) The biodiversity crisis: a challenge for biology. *Oikos* 63: 29–38. doi: 10.2307/3545513
- Wheeler QD, Raven PH, Wilson EO (2004) Taxonomy: Impediment or Expedient? *Science* 303: 285. doi: 10.1126/science.303.5656.285
- Wiens JJ, Penkrot TA (2002) Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). *Systematic Biology* 51: 69–91. doi: 10.1080/106351502753475880
- Wilson EO (1988) *Biodiversity*. National Academy Press, Washington, DC, 538 pp.
- Yang Z (1998) On the best evolutionary rate for phylogenetic analysis. *Systematic Biology* 47: 125–133. doi: 10.1080/106351598261067
- Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences USA* 107: 9264–9269. doi: 10.1073/pnas.0913022107
- Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3: 613–623. doi: 10.1111/j.2041-210X.2012.00198.x
- Zhang JJ, Kapli P, Pavlidis P, Stamatakis AA (2013) General species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29: 2869–2876. doi: 10.1093/bioinformatics/btt499
- Zhou X, Li YY, Liu SL, Yang Q, Su X, Zhou LL, Tang M, Fu RB, Li JG, Huang QF (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience* 2: 4. doi: 10.1186/2047-217X-2-4